

# Chapter 15

## A Dynamic Modeling and Validation Framework for the Market Direction Prediction

**Xugang Ye**

*Johns Hopkins University, USA*

**Jingjing Li**

*University of Colorado, USA*

### ABSTRACT

*There are many research papers talking about building various machine learning models to predict the market index. However, very few attention has been paid to effectively validating or calibrating the prediction results. The focus of this paper is to present a dynamic modeling and validation framework for the market direction prediction. The central idea is to calibrate the probabilistic prediction by estimating two conditional probabilities of correct forecast from the dynamic validation data set. The calibration method can be combined with any predictive model that generates probabilistic prediction of the market direction.*

### 1. INTRODUCTION

Modeling and forecasting market direction to facilitate trading strategies have been receiving more and more attention from academia and industry (Christoffersen, 2006; Granger, 1992; Tsay, 2010). Aiming at forecasting price movement — up or down — of general stock market, market direction prediction is also called “market timing” (Henriksson & Merton, 1981). Undoubtedly, this task is difficult because of the high market volatility. Many factors, such as political event, economic condition, trade mismatch, rumors, news and investors’ sentiment and mentality, all contribute to the high degree of uncertainty and fluctuation of the stock market (Harris, 2008).

Numerous efforts have been dedicated to find adequate modeling techniques to capture the market volatility. For example, the time series models, such as Autoregressive Conditional Heteroscedasticity (ARCH) (Engle, 1982) and generalized ARCH (GARCH) (Bollerslev, 1986), have been extensively

DOI: 10.4018/978-1-5225-1759-7.ch015

used in economic and finance research. Those models, originated from the theories of the financial time series, often assume a stationary linear correlation structure (Tsay, 2010) among the time series data, which may not be able to capture the non-stationary nonlinear patterns and the impact of external events. On the contrary, machine learning approaches focus on finding patterns from data and usually make much less assumptions. Although various types of machine learning models including Artificial Neural Network (ANN) (Nicholas Refenes et al., 1994; Yoon & Swales, 1991; Zhang & Wu, 2009), Support Vector Machine (SVM) (Cao & Tay, 2003; Huang et al., 2005; Rao & Hong, 2010), Adaboosting (Rodríguez & Sosvilla-Rivero, 2006), and Hidden Markov Model (HMM) (Hassan & Nath, 2005; Rao & Hong, 2010) have been applied to the area of stock market forecasting, the machine learning methods became dominant only recently due to the fact that the technologies that enable people to generate and handle massive data are only recent stories.

Over the past few years, building machine learning models to predict financial markets has become more and more popular (Atsalakis & Valavanis 2009; Yoo et al., 2005). However, the existing market prediction literature shows that most work centers on static models, that is to build a model under a variety of parameter settings from a training data set of a long time window, pick the best one according to its performance on the validation data set (as part of the training data set), and then test it on a hold-out data set of a relatively short time window for evaluation. Although this is a sound procedure for many machine learning applications, for market prediction, a problem is that there could be several market regime changes during a long time horizon so that a model trained and validated from the data collected from a long historical time window may not be able to catch the market dynamics. Another problem is that the models/parameters that are thrown away by the one-time validation will never participate the future recursive prediction tasks. Consequently, the market prediction is solely dependent on a narrow range of model settings, which are hardly adaptive to the future market changes. One fact that is often neglected is that for market prediction, a model that has very poor validation performance does not necessarily mean it should be discarded, one can use the opposite of the prediction. It actually follows that a really bad model is the one that has equal chance of predicting right or wrong. We say this kind of models do not generate “signal” and performs as same as random guess. Likewise, a model that performs very well on the validation set does not necessarily do as well in continual tests. One simple reason is that the market is driven by much more factors than the finite number of features that any model can accommodate. Tightly sticking to a model will likely lead to catastrophe sometime in the future (Harris, 2008).

Our remedy is to combine the rolling procedure (Tsay, 2010) that is widely used in time series analysis and a calibration step that is derived from the old results of Henriksson and Merton (1981). In short, we build machine learning models from the data collected from a rolling time window to predict the daily directional move of the market. As the window slides forward, a validation set can be formed as long as there are enough observed targets. And when the window moves further, the validation set updates by absorbing a newly observed target and spitting out an old one. Dynamically, we calibrate the prediction generated by any machine learning model according to the validation performance. Depending on the validation results, we decide whether to stick to the prediction, use the opposite of the prediction, or follow the market — all of them could be reflected in the calibrated prediction.

The rest sections of this paper are organized as follows: we first introduce four machine learning models — logistic regression, linear regression, fisher’s linear discriminant analysis and SVM with soft margin — that are used in our work to predict the market direction. We then present the principal component

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-dynamic-modeling-and-validation-framework-for-the-market-direction-prediction/173343](http://www.igi-global.com/chapter/a-dynamic-modeling-and-validation-framework-for-the-market-direction-prediction/173343)

## Related Content

---

### **Imprecise Knowledge and Fuzzy Modeling in Materials Domain**

Subhas Ganguly and Shubhabrata Datta (2017). *Fuzzy Systems: Concepts, Methodologies, Tools, and Applications* (pp. 170-183).

[www.irma-international.org/chapter/imprecise-knowledge-and-fuzzy-modeling-in-materials-domain/178393](http://www.irma-international.org/chapter/imprecise-knowledge-and-fuzzy-modeling-in-materials-domain/178393)

### **Towards an Automated Model to Evaluate Collaboration Through Non-Verbal Interaction in Collaborative Virtual Environments**

Luis Casillas, Adriana Peña and Alfredo Gutierrez (2018). *Intelligent Systems: Concepts, Methodologies, Tools, and Applications* (pp. 1570-1586).

[www.irma-international.org/chapter/towards-an-automated-model-to-evaluate-collaboration-through-non-verbal-interaction-in-collaborative-virtual-environments/205847](http://www.irma-international.org/chapter/towards-an-automated-model-to-evaluate-collaboration-through-non-verbal-interaction-in-collaborative-virtual-environments/205847)

### **Bridging the Gap Between Narrative Generation Systems and Narrative Contents With Kabuki-Oriented Narratology and Watakushi Monogatari**

Takashi Ogata (2021). *Bridging the Gap Between AI, Cognitive Science, and Narratology With Narrative Generation* (pp. 126-248).

[www.irma-international.org/chapter/bridging-the-gap-between-narrative-generation-systems-and-narrative-contents-with-kabuki-oriented-narratology-and-watakushi-monogatari/261702](http://www.irma-international.org/chapter/bridging-the-gap-between-narrative-generation-systems-and-narrative-contents-with-kabuki-oriented-narratology-and-watakushi-monogatari/261702)

### **A Particle Swarm Optimization Algorithm for Web Information Retrieval: A Novel Approach**

Tarek Alloui, Imane Boussebough and Allaoua Chaoui (2015). *International Journal of Intelligent Information Technologies* (pp. 15-29).

[www.irma-international.org/article/a-particle-swarm-optimization-algorithm-for-web-information-retrieval/139468](http://www.irma-international.org/article/a-particle-swarm-optimization-algorithm-for-web-information-retrieval/139468)

### **A New Technique to Solve Transportation Problems Using De-Pythagorean Value in Generalized Interval Valued Pythagorean Environment**

Avishek Chakraborty, Soheil Salahshour and Tofigh Allahviranloo (2021). *International Journal of Fuzzy System Applications* (pp. 57-78).

[www.irma-international.org/article/a-new-technique-to-solve-transportation-problems-using-de-pythagorean-value-in-generalized-interval-valued-pythagorean-environment/288395](http://www.irma-international.org/article/a-new-technique-to-solve-transportation-problems-using-de-pythagorean-value-in-generalized-interval-valued-pythagorean-environment/288395)